Appendix

A Functions

We list the functions considered in this work in Table 1. Each function can only be applied on certain kinds of nodes, i.e., its domain.

B Query Minimization

Here we give a more formal description of our query minimization algorithm. Suppose a knowledge base $\mathcal{K}$ is given to compute the answer of graph queries, then the following concepts can be naturally defined.

Definition 1 (Redundant Component) *A component in a graph query q is* redundant *iff. removing it does not change the answer $[\![q]\!]_\mathcal{K}$.*

Definition 2 (Minimal Query) *A graph query q is* minimal *iff. there is no sub-query q', i.e., a graph query resulted from removing any number of components from q, such that $[\![q']\!]_\mathcal{K} = [\![q]\!]_\mathcal{K}$, except q itself.*

Definition 3 (Equivalent Minimal Query) *Given two graph queries q and q', q' is an* equivalent minimal query *of q iff. (1) $[\![q']\!]_\mathcal{K} = [\![q]\!]_\mathcal{K}$, and (2) q' is minimal.*

Note that we define redundancy based on a given knowledge base because it suffices for our purpose. KB-independent minimization is out of the scope of this work, but is of interest for future study.

Our query minimization algorithm is outlined in Algorithm 1. We first examine every edge (in an arbitrary order), and remove an edge if it is redundant. So the `dateOfBirth` edge in Figure 1(a) will be removed. Some nodes may become disconnected to the question node after removing redundant edges, which indicates that the node is redundant. We thus delete the redundant nodes as well, e.g., the date node in Figure 1(a). The query in Figure 1(b) is produced as output.

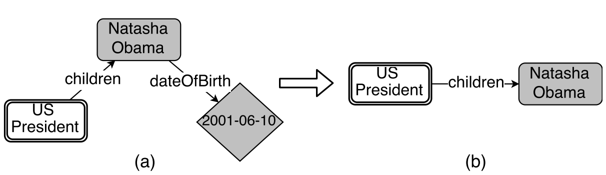


Figure 1: Query minimization: (a) a query with redundant components, (b) a corresponding equivalent minimal query.

Proof. The proof is straightforward. For convenience, we assume there is only one question node, but the proof generalizes to graph queries with multiple question nodes easily. When there is no function on the question node (`count`, `max` or `min`), the answer is a set of individuals. Since a component in a graph query imposes some constraint on the answer, removing it will not eliminate any individual from the answer, rather, the answer will only stay unchanged or get expanded. More formally, given a knowledge base $\mathcal{K}$, for a graph query q and a sub-query q', we have $[\![q]\!]_\mathcal{K} \subseteq [\![q']\!]_\mathcal{K}$. When there is a function on the question node, the answer is a single number, but the same assertion holds. For example, if there is a `count` function on the question node, it is asserted that the answer, i.e., the answer cardinality of the graph query without the function, will be non-decreasing when iteratively removing other components of the graph query. □

Algorithm 1: Graph Query Minimization

Input: a graph query q, a knowledge base $\mathcal{K}$
Output: an equivalent minimal query of q
foreach *edge e of q* **do**
  $q' \leftarrow q$ with e removed
  if $[\![q']\!]_\mathcal{K} = [\![q]\!]_\mathcal{K}$ **then**
    $q \leftarrow q'$
  end
end
Remove nodes disconnected to the question node
return q

To prove the generated query, denoted as q', is an equivalent minimal query of the input query q, we need to prove (1) $[\![q']\!]_\mathcal{K} = [\![q]\!]_\mathcal{K}$, and (2) q' is minimal. The former is guaranteed by the construction of Algorithm 1. To prove the latter, we need the following lemma:

Lemma 1. *When components (except for the question nodes) are iteratively removed from a graph query, the answer will change monotonically.*

Lemma 1 precludes such possibilities: Removing one component changes the answer, and subsequently removing another component changes the

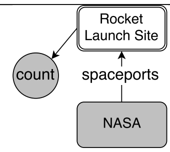
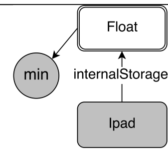
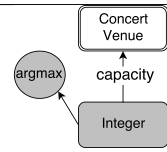
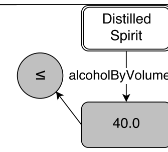
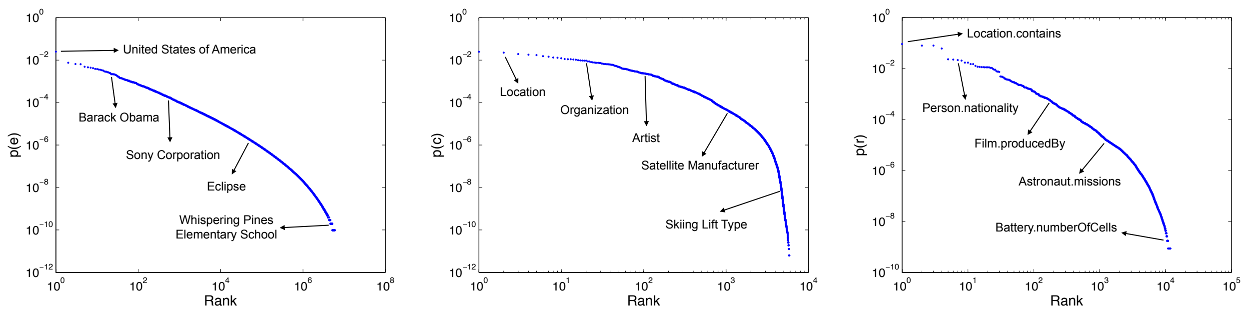
Category	Counting	Superlative		Comparative
Functions	`count`	`max` and `min`	`argmax` and `argmin`	$<$,$>$,$\leq$, and $\geq$
Domain	Question node	Question node of numeric class	Template/grounded node of numeric class	Template/grounded node of numeric class
Example				
Question	How many launch sites does nasa have?	What's the smallest internal storage of ipad?	Find the largest concert venue.	List distilled spirits with no more than 40.0% abv.

Table 1: Functions considered in this work. Domain is the type of nodes a function can be applied on.



(a) entity (b) class (c) relation

Figure 2: Component probability distributions. Components are ranked in descending order by their probability.

answer back. Therefore, it is not necessary to enumerate all the sub-queries in order to check the minimality of a graph query. Rather, we only need to examine the redundancy of each component. In other words, we have the following corollary:

Corollary 1.1. *A graph query is minimal iff. it has no redundant component.*

In addition, a single scan of the edges in an arbitrary order, as Algorithm 1 does, is sufficient, because an irredundant component will not become redundant when other components are removed.

Theorem 2. *The graph query resulted from Algorithm 1 is an equivalent minimal query of the input query.*

Proof. The answer stays unchanged by construction. After removing the redundant edges and nodes, the resulted query has no redundant component, and is therefore minimal according to Corollary 1.1. Following the definition, the resulted query is an equivalent minimal query of the input query. □

C Probability Distribution of Freebase Components

We count the frequency n' of each component in the Freebase. For an entity e, $n'(e) = 1$; for a class c, $n'(c)$ is its number of instances; and for a relation r, $n'(r)$ is the number of facts of this relation. We then add the mention counts from FACC1 with the counts from Freebase to estimate the probability of Freebase components. The probability of literal classes are solely determined by their number of instances, and the probability of a literal instance is the same as the corresponding literal class. The distributions are shown in Figure 2. Entities with no mention in FACC1 are not shown in Figure 2(a). Classes and relations are from a filtered ontology where the `User` domain and the `Freebase` domain are removed. Relations shown in the figure are in the *class.relation* format.

D Fine-grained Statistics of GraphQuestions

Fine-grained statistics of GraphQuestions on different characteristics are shown in Figure (3-9).

Sub-distributions of the training set and the testing set are also shown. Note that in addition to the commonness of the whole query (Figure 6), we also give the commonness of topic entities (Figure 7). As a comparison, we also give the commonness distribution of topic entities of WEBQUESTIONS (Figure 8). GRAPHQUESTIONS contains topic entities over a broader range of commonness than WEBQUESTIONS. Overall GRAPHQUESTIONS exhibits a good diversity in all the examined characteristics. More example questions are listed in Table 2.

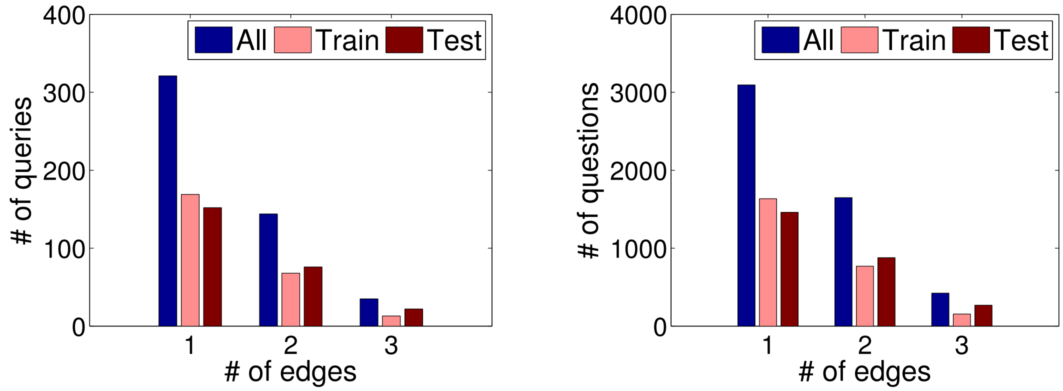
E Experiment Configuration Details

All the systems (SEMPRE, PARASEMPRE, and JACANA) are trained on the training set and tested on the testing set of GRAPHQUESTIONS, and use Freebase as the knowledge base. For SEMPRE, the grammar is the one from the original paper, the maximum training iteration is 3, the beam size is 100 for training, 200 for testing[1]. The training took 5 days. For PARASEMPRE, the maximum training iteration is 3, the number of threads is 20 for training, 1 for testing. The training took 27 hours. Both SEMPRE and PARASEMPRE cache historical SPARQL query results in order to save time. The cache from training is allowed to use during testing. For JACANA, the top-1 topic entity retrieved from the Freebase Search API is used for both training and testing[2]. Same as the original paper, we down-sample the negative examples with a ratio of 0.2. The classifier is logistic regression with L1 regularization. The training took 1.5 hours.

Experiments are run on a Linux server with Intel Xeon E7-8837 processors (2.67GHz) and 1T memory. A Virtuoso database is run on the same server to provide access to Freebase, so network IO cost is minimized.

[1] It is too time consuming to use beam size=200 for training

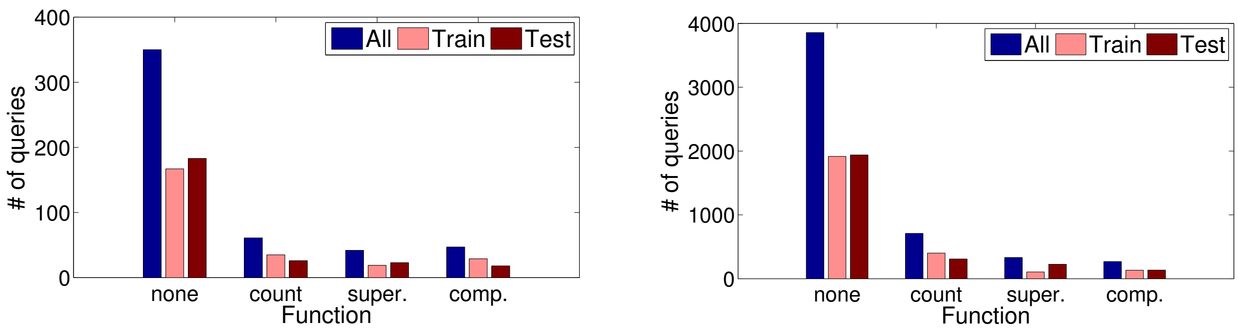
[2] We also tried to use the top-10 retrieved topic entities for both training and testing, or use the *gold* topic entities for training and the top-10 retrieved topic entities for testing. The performance was slightly worse.



(a) Structure distribution of graph queries

(b) Structure distribution of questions

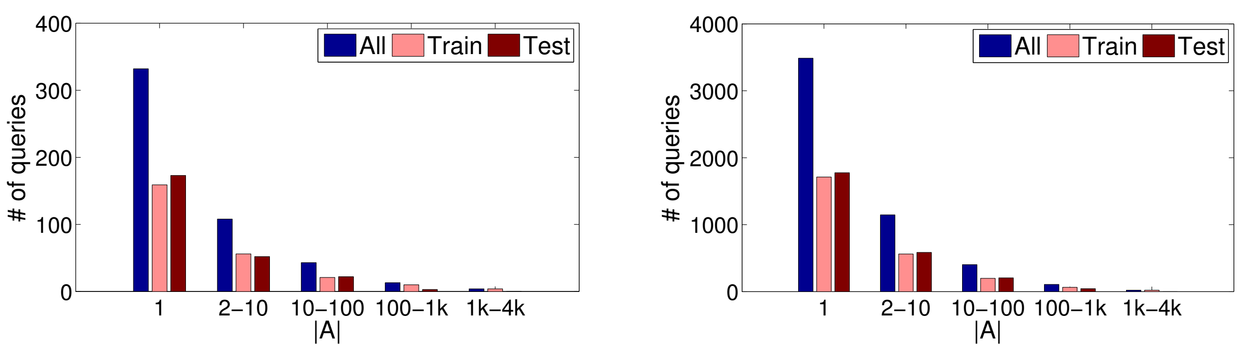
Figure 3: Characteristics distribution of GRAPHQUESTIONS: Structure Complexity.



(a) Function distribution of graph queries

(b) Function distribution of questions

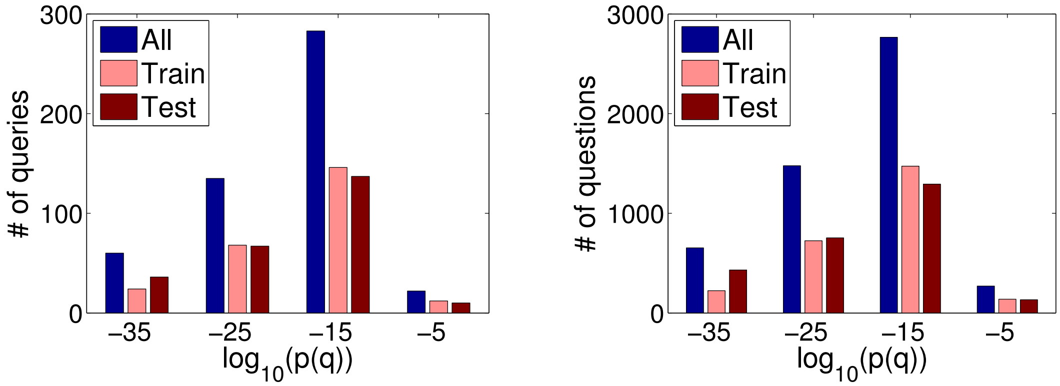
Figure 4: Characteristics distribution of GRAPHQUESTIONS: Function.



(a) Answer cardinality distribution of graph queries

(b) Answer cardinality distribution of questions

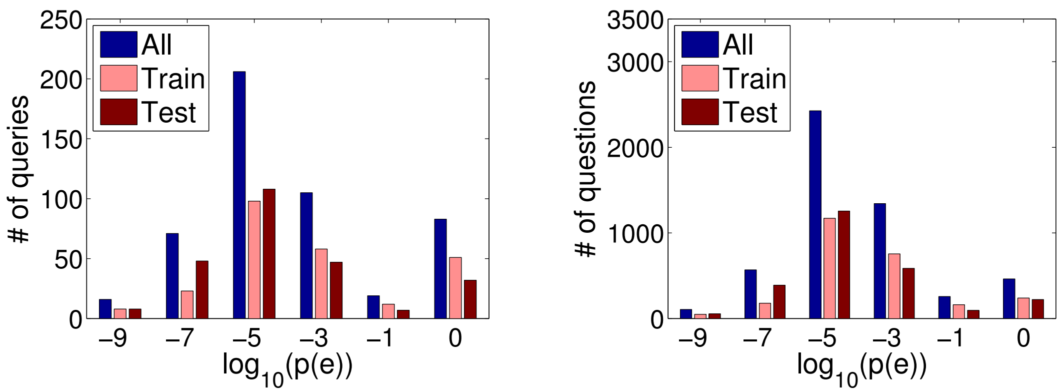
Figure 5: Characteristics distribution of GRAPHQUESTIONS: Answer Cardinality.



(a) Commonness distribution of graph queries

(b) Commonness distribution of questions

Figure 6: Characteristics distribution of GRAPHQUESTIONS: Commonness. Note that $x = -5$ indicates the commonness range $-10 \leq \log_{10}(p(q)) < 0$.



(a) Topic commonness distribution of graph queries

(b) Topic commonness distribution of questions

Figure 7: Characteristics distribution of GRAPHQUESTIONS: Topic Commonness. When there are multiple topic entities, the most common one is used. Note that $x = -5$ indicates the commonness range $-6 \leq \log_{10}(p(q)) < -4$. $\log_{10}(p(e)) = 0$ means there is no topic entity, e.g., "*What's the smallest Spanish autonomous city?*"

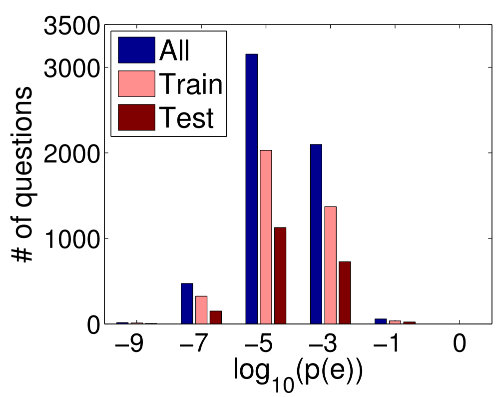
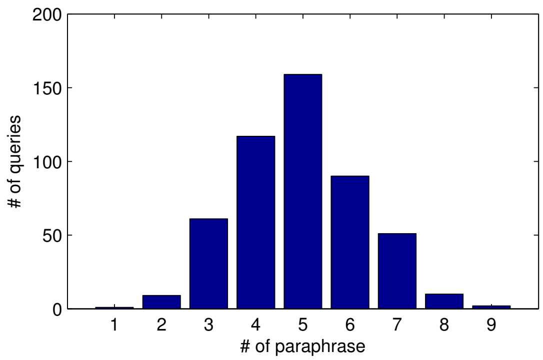
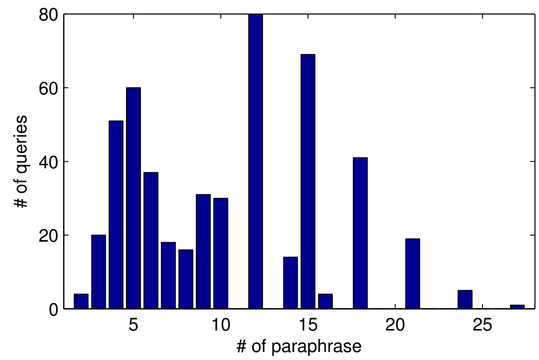


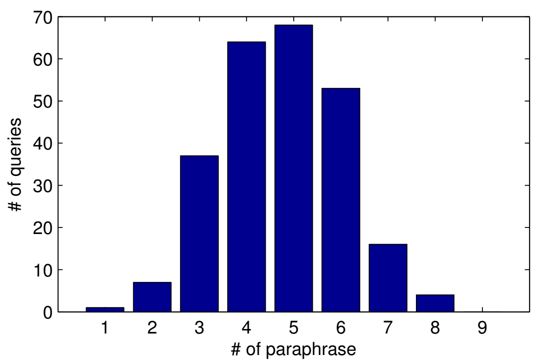
Figure 8: Topic commonness distribution of WEBQUESTIONS. Note that $x = -5$ indicates the commonness range $-6 \leq \log_{10}(p(q)) < -4$. $\log_{10}(p(e)) = 0$ means there is no topic entity, e.g., "*What's the smallest Spanish autonomous city?*"



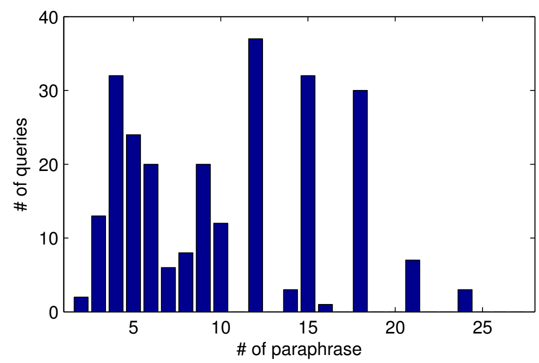
(a) Sentence paraphrase distribution of all queries



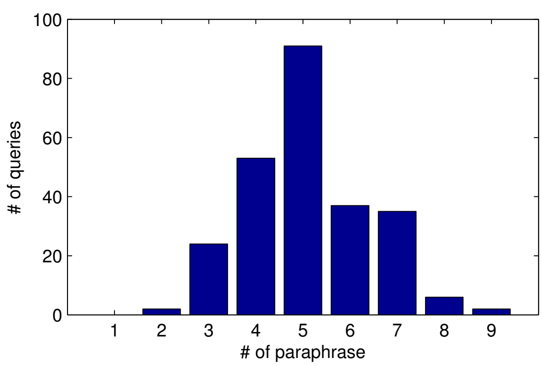
(b) Sentence+Entity paraphrase distribution of all queries



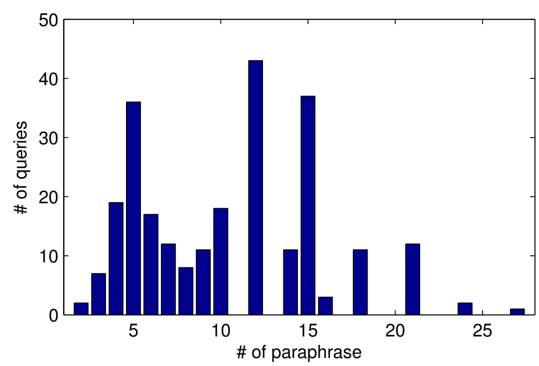
(c) Sentence paraphrase distribution of training queries



(d) Sentence+Entity paraphrase distribution of training queries



(e) Sentence paraphrase distribution of testing queries



(f) Sentence+Entity paraphrase distribution of testing queries

Figure 9: Characteristics distribution of GRAPHQUESTIONS: Paraphrase. The x-axis is the number of paraphrases, while the y-axis indicates how many graph queries have this number of paraphrases.

Question	Domain	Answer	# of edges	Function	$\log_{10}(p(q))$	$\lvert\mathbf{A}\rvert$
how many scottish clans live in **united kingdom**? give me the count of scottish clans in **united kingdom**. how many clans from scotland still exist in the **uk**?	Scottish Clans	4	1	count	-12.11	1
what solid materials fuse at **435.4** joules/mole and above? which materials need to take at least **435.4** joules of heat to fuse for one mole? name all the solid materials the fusion of which need at least **435.4** joules for a mole.	Materials	Titanium, Beryllium, Aluminium oxide	1	comp.	-17.04	3
who influenced **paul the apostle**? by whom was **paul the apostle** influenced? who made a significant influence on **paul**?	Influence	Jesus Christ	1	none	-8.93	1
find rockets made by **chrysler group llc** that support **low earth orbit**. which of **chrysler group llc**'s rockets are capable of **low earth orbit**? which **low earth orbit** rockets are made by **chrysler**?	Spaceflight	Saturn I, Saturn IB	2	none	-25.76	2
what is the nutritional composition of **coca-cola** soda? what is the supplement information for **coca-cola**? what kind of nutrient does **coke** have?	Food	Sugar, Caffeine . . .	2	none	-18.34	19
which tropical cyclone in the **2008 atlantic hurricane season** caused the most fatalities? which **2008 atlantic hurricane season**'s tropical cyclone was the most deadly? which of the **atlantic hurricane season 2008**'s tropical cyclones killed the most people?	Meteorology	Hurricane Hanna	2	super.	-29.36	1
people who are on a **gluten-free diet** can't eat what cereal grain that is used to make **challah**? which cereal grain which can be utilized for making **challah** is unable to be consumed by those on a **gluten-free diet**? what cereal grain can be used to produce **challah**, and people on **gluten free** could not eat?	Food	Wheat	3	none	-39.48	1
what's the theme of the casino having the most rooms under the control of **caesars entertainment corporation**? what type of theme is **caesars entertainment corporation**'s largest casino made of? how is the largest casino owned by **harrahs** themed?	Casinos	Art Deco	3	super.	-34.11	1

Table 2: More example questions and characteristics. Topic entities are bold-faced. Three sentence-level paraphrases are shown for each graph query, with entity level paraphrasing applied on the third. Questions are lowercased.